Biosystematics and Ecology 1: e71620
doi: 10.1553/biosystecol.1.e71620

OPEN ACCESS

Research Article

# Semantic-based methods for morphological descriptions: An applied example for Neotropical species of genus *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae)

Luis Antonio González-Montaña ‡

‡ Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Bogotá, Colombia

Corresponding author: Luis Antonio González-Montaña (lagonzalezmo@unal.edu.co)

Academic editor: Christian Sturmbauer

Received: 15 Jul 2021 | Accepted: 28 Sep 2021 | Published: 11 Nov 2021

Citation: González-Montaña LA (2021) Semantic-based methods for morphological descriptions: An applied example for Neotropical species of genus *Lepidocyrtus* Bourlet, 1839 (Collembola: Entomobryidae). Biosystematics and Ecology 1: e71620. https://doi.org/10.1553/biosystecol.1.e71620

## Abstract

The production of semantic annotations has gained renewed attention due to the development of anatomical ontologies and the documentation of morphological data. Two methods are proposed in this production, differing in their methodological and philosophical approaches: class-based method and instance-based method. The first, the semantic annotations are established as class expressions, while in the second, the annotations incorporate individuals. An empirical evaluation of the above methods was applied in the morphological description of Neotropical species of the genus *Lepidocyrtus* (Collembola: Entomobryidae: Lepidocyrtinae). The semantic annotations are expressed as RDF triple, which is a language most flexible than the Entity-Quality syntax used commonly in the description of phenotypes. The morphological descriptions were built in Protégé 5.4.0 and stored in an RDF store created with Fuseki Jena. The semantic annotations based on RDF triple increase the interoperability and integration of data from diverse sources, e.g., museum data. However, computational challenges are present, which are related with the development of semi-automatic methods for the generation of RDF triple, interchanging between texts and RDF triple, and the access by non-expert users.

© González-Montaña L. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Keywords

class-based methods, instance-based methods, resource description framework (RDF), Neotropical region, Hexapoda

## Introduction

In recent years, anatomical ontologies have gained attention in the formalization of semantic-based morphological descriptions, which encompass the standardization of anatomical terminology and interoperability between morphological data coming from diverse sources (Dahdul et al. 2015). This formalization implies new computational tools and common language that allows building of semantic annotations parsable by a non-human reader. A review of formalized language and software developed for semantic annotations was realized by Yoder et al. (2018). Three standardized languages are employed by the World Wide Web Consortium (W3C): 1) XML together with the introduction of CharaParser (Cui 2012, Cui et al. 2016), 2) NeXML linked to phenotypic descriptions in OWL (Balhoff et al. 2013), and 3) Resource Description Framework (RDF) (Vogt 2017). Software most used for semantic annotations is Phenex, a platform that employs the Entity-Quality syntax (EQ model), which could be translated to character statement (Balhoff et al. 2010):

chaeta (Entity): ciliated (Quality)

chaeta (character): ciliated (character state)

This software was based initially in NeXML language, incorporating after mx, a web-based application to gather information about specimens, and building descriptive matrices (Balhoff et al. 2013). Nevertheless, NeXML is challenged for non-expert users and mx is disabled currently. On the other hand, the class-based method presupposes the identification *a priori* of comparative homologs (Vogt 2017). The EQ model is a class-based method, where the generalizations and phenotypes are described within the definition of ontology class (Vogt 2019), containing TBox assertions, axioms that defining classes, and the relations between them (De Giacomo and Lenzerini 1996). This model is employed in morphological descriptions and building of character statements (Balhoff et al. 2012, Burks et al. 2016, Trietsch et al. 2018, Yoder et al. 2018, Tarasov 2019)

A second approach called “semantic instance anatomy” is developed by Vogt (2017) and Vogt (2019). In this method, each anatomical entity and its qualities are represented by individual resources, being themselves instances (Vogt 2019), containing ABox assertions, axioms on individual objects or instances (De Giacomo and Lenzerini 1996). While the class-based method contains metadata from different organisms, “semantic instance anatomy” contains metadata for an organism documented in a separate semantic graph (Vogt 2019). The documentation is made through ontology classes from Ontology for Biomedical Investigations (OBI) to specify how the specimens are preserved, and from Biological Collection Ontology to related collection information as catalog number or

collection number (Vogt 2019). Both methods employ predicates that describe topological relations between anatomical entities, and relations between anatomical entities and their qualities.

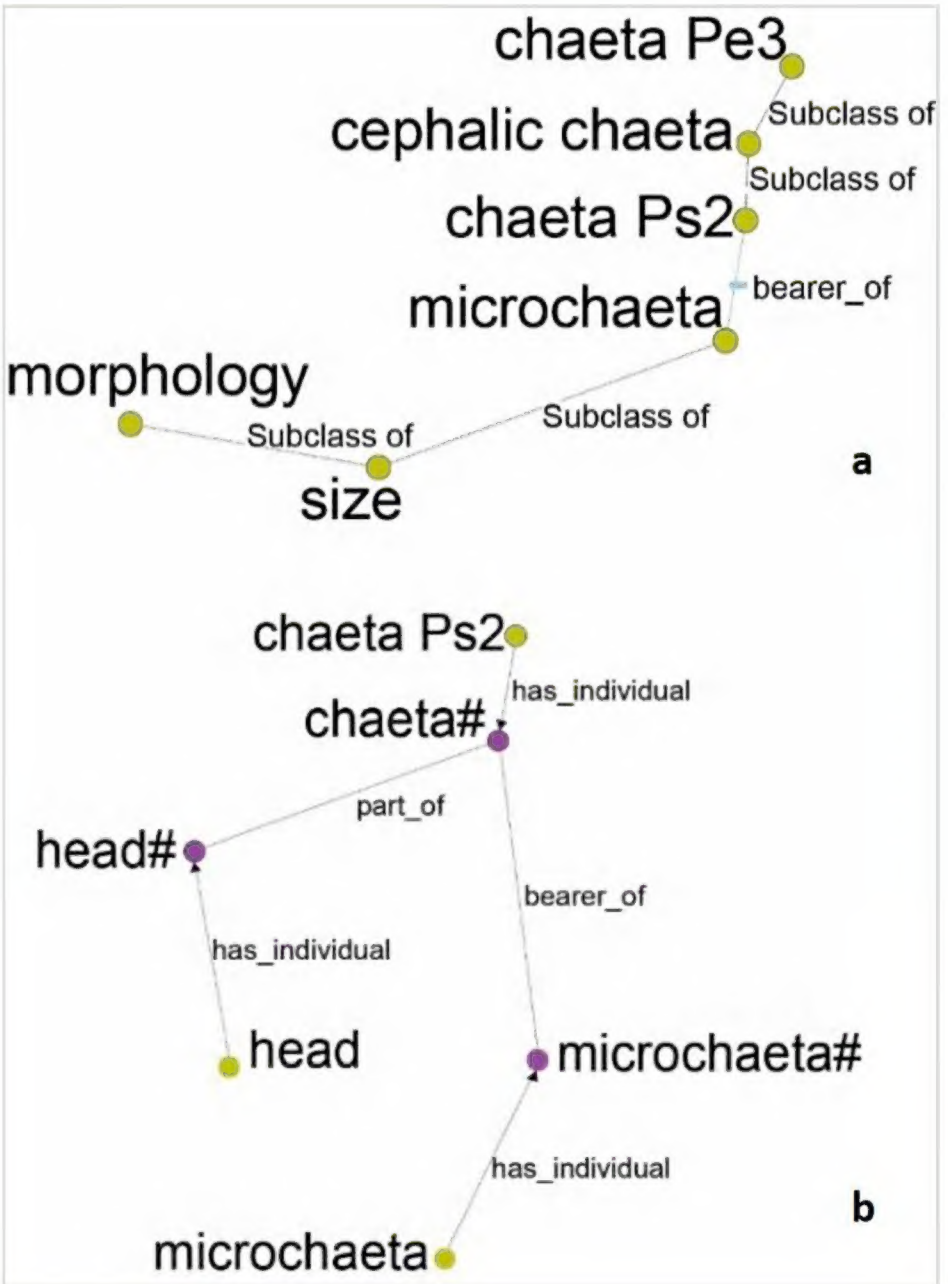


Figure 1. doi

Graphs of internal structure of **a**) class-based and **b**) instance-based methods. The orange points represent classes within the descriptive template that in the example refers to the chaeta Ps2 part of cephalic chaetae. The purple points represent instances or individuals for each class and are named with provisional labels. The SubClassOf relation links classes, and the has_individual relation links class and individual instanced.

Philosophical differences rely on objects that are described in each method, classes, and instances (Fig. 1), and how these are approached epistemologically. While the classes are generalizations, the instances are the result of direct observation of objects (parts of organisms) that exist in the reality and independent of human cognition (Mahner and Bunge 1997). The development of semantic-based morphological descriptions has been extensive within Hymenoptera (Mikó et al. 2015, Mikó et al. 2016, Silva and Feitosa 2019). Of course, the application of these semantic methods is strongly related to the building of anatomical ontologies, but these are scarce and only available for few taxa, Hymenoptera

Ontology Anatomy (HAO) (Yoder et al. 2010), Mosquito Gross Anatomy Ontology (TGMA) (Grumbling and Strelets 2006), Drosophila Gross Anatomy (FBBT) (Grumbling and Strelets 2006), Spider Ontology (SPD) (Ramírez and Michalik 2019), Tick Anatomy Ontology (TADS) (Topalis et al. 2008), and Ontology for the Anatomy of the Insect SkeletoMuscular system (AISM) (Girón et al. 2021). Recently, Collembola Anatomy Ontology (CLAO) (Gonzalez-Montaña 2021) has been developed (available at www.ontobee.org), offering an opportunity in the application of semantic methods for morphological descriptions in Collembola. The goal of this paper is a demonstration of the applicability of semantic-based methods in morphological descriptions for the Neotropical species of genus *Lepidocyrtus* Bourlet, 1839, the second larger genus within Entomobryidae, with almost 512 species worldwide (Bellinger et al. 2021) and 40 species described for the Neotropical Region (Mari Mutt and Bellinger 1990).

# Material and methods

## Descriptive templates

Unfortunately, approaches to the production of semantic annotations in a standardized language are scarce or not available for non-expert users. A method is the building of semantic spreadsheet templates where classes, individuals, and properties are declared, with which an ontology is built. A good example is developed by Girón et al. 2021 found at https://github.com/insect-morphology/aism/blob/master/AISM_template_examples.tsv, where the description of cercus through axioms is made for various insects orders. Another application is proto.morphdbase.de, which allows the declaration of individuals, but currently not available. In this paper, the morphological descriptions are based on the building of a "descriptive template", which is a .owl file to hold the anatomical terms (e. g. chaetotaxy) manually imported from CLAO. The import involves the search and the declaration within the template of terms related with anatomical entities described for *Lepidocyrtus.* Optional tools to import terms could be useful as ODK (https://github.com/INCATools/ontology-development-kit/) or Ontofox (http://ontofox.hegroup.org), increasing the performance in this step, but the terms curation must be done by an expert in the taxon. The "descriptive template" has classes and subclasses organized in a hierarchical structure, where class expressions or direct instances are the semantic annotations for each species. For instance, class expression for the absence of scales on antennae is: "**antenna** *SubClassOf* not (*bearer_of* some **scaled**)", while the presence of scales is expressed as "**antenna** *SubClassOf* (*bearer_of* some **scaled**)".

Continuing with the example above, the class **antenna** represents an individuals set that shares some property, e. g., part of the head capsule. The individuals are described object property or data property assertion could be made as:

"**antenna** *has_individual* some antenna123", where the individual labeled as "antenna123" is the instance of the class **antenna**.

and

"antenna123 *part_of* headcapsule123", where "*part_of* abdominalsegment1" is the object property assertion and "headcapsule123" is a provisional label for an individual that is an instance of the class **head capsule**.

Under the class-based method, the template has a "layer", the class expressions, and under the instance-based method, the template has two "layers", the class expressions, and individuals by class.

The morphological descriptions follow the HYA model proposed by Wirkner et al. 2017, including the phenotypic categories: 1. qualitative phenotypes, 2. presence/absence phenotypes, 3. count phenotypes, and 4. relative measurement phenotypes. Terminology related to the description of phenotypes were taken from Phenotype And Trait Ontology (PATO) (Table 1): anatomical side, color pattern, length, morphology (shape, size, and texture), and number; the spatial object properties are enlisted in the Table 2. Semantic annotations are declared through OWL class expressions and Manchester Syntax (http://www.w3.org/TR/owl2-manchester-syntax/) and built in Protégé 5.4.0 Musen 2015.

Table 1.

Predicates or relations employed during the expression of RDF triple statement under class-based and instance-based methods. Relation Ontology (RO); Phenotype And Trait Ontology (PATO).

| Term | definition | Ontology identifier |
|---|---|---|
| globular | A spheroid quality inhering in a bearer by virtue of the bearer's resembling a ball | PATO_0001499 |
| cylindrical | A convex 3-D shape quality inhering in a bearer by virtue of the bearer's exhibiting a consistently sized round cross section | PATO_0001873 |
| absent | A quality denoting the lack of an entity. | PATO_0000462 |
| rounded | A shape quality inhering in a bearer by virtue of the bearer's being such that every part of the surface or the circumference is equidistant from the center | PATO_0000411 |
| Protruding | A quality inhering in a bearer by virtue of the bearer's extending out above or beyond a surface or boundary. | PATO_0001598 |
| truncated | A shape quality inhering in a bearer by virtue of the bearer's terminating abruptly by having or as if having an end or point cut off | PATO_0000936 |
| lanceolate | A shape quality inhering in a bearer by virtue of the bearer's being shaped like a lance-head, considerably longer than wide, tapering towards the tip from below the middle; attached at the broad end | PATO_0001877 |
| serrated | A shape quality inhering in a bearer by virtue of having sharp straight-edged teeth pointing to the apex. | PATO_0001206 |
| asymmetrically curved | A curvature quality inhering in a bearer by virtue of the bearer's being curved asymmetrically. | PATO_0001848 |

| Term | definition | Ontology identifier |
|---|---|---|
| domed | A curvature quality inhering in a bearer by virtue of the bearer's having a shape resembling a dome. | PATO_0001789 |
| increased curvature | A curvature which is relatively high. | PATO_0001592 |
| distributed | A spatial pattern inhering in a bearer by virtue of the bearer's being spread out or scattered about or divided up. | PATO_0001566 |
| subulate | A shape quality inhering in a bearer by virtue of the bearer's being linear, very narrow, tapering to a very fine point from a narrow base. | PATO_0001954 |
| filamentous | A shape quality inhering in a bearer by virtue of the bearer's having thin filamentous extensions at its edge. | PATO_0001360 |
| semicircular | A 2-D shape quality inhering in a bearer by virtue of the bearer's having shape or form of half a circle. | PATO_0002232 |
| branched | A branchiness quality inhering in a bearer by virtue of the bearer's having branches. | PATO_0000402 |
| increased amount | An amount which is relatively high. | PATO_0000470 |
| decreased amount | An amount which is relatively low. | PATO_0001997 |
| right side of | A spatial quality inhering in a bearer by virtue of the bearer's being located on right side of a another entity. | PATO_0001793 |
| left side of | A spatial quality inhering in a bearer by virtue of the bearer's being located on left side of from the a another entity. | PATO_0001792 |
| aligned with | An alignment quality inhering in a bearer by virtue of the bearer's being in a proper spatial positioning with respect to an additional entity. | PATO_0001653 |
| normal | A quality inhering in a bearer by virtue of the bearer's exhibiting no deviation from normal or average. | PATO_0000461 |

## RDF repository

An RDF store was built with Apache Jena Fuseki, an HTTP interface for querying RDF graphs, which can be explored in a browser as http://localhost:3030//query.html, employing SPARQL according to W3C recommendations (The World Wide Web Consortium). Apache Jena Fuseki was chosen by the simplicity in the installation, in contrast with another web API with SPARQL endpoint, for instance, Openlink Virtuoso, Ontotext, or Neo4j. This created RDF store only holds class-based morphological descriptions for each described species.

Table 2.

Predicates or relations employed during the expression of RDF triple statement under class-based and instance-based methods. Relation Ontology (RO); Phenotype And Trait Ontology (PATO). The relation *bearer_of* is a aternative term of *has_characteristic* (RO:0000053), which could be available.

| Imported relation property | Ontology | PURL |
|---|---|---|
| has_part | RO | http://purl.obolibrary.org/obo/BFO_0000051 |
| adjacent_to | RO | http://purl.obolibrary.org/obo/RO_0002220 |
| aligned_with | RO | http://purl.obolibrary.org/obo/RO_0002001 |
| anterior_to | PATO | http://purl.obolibrary.org/obo/PATO_0001632 |
| bearer_of | RO | http://purl.obolibrary.org/obo/RO_0000053 |
| decreased_in_magnitude_relative_to | PATO | http://purl.obolibrary.org/obo/pato #decreased_in_magnitude_relative_to |
| external_to | RO | http://purl.obolibrary.org/obo/PATO_0002483 |
| increased_in_magnitude_relative_to | PATO | http://purl.obolibrary.org/obo/pato #increased_in_magnitude_relative_to |
| internal_to | Not available | |
| is_approximately_equivalent_to | RO | http://purl.obolibrary.org/obo/RO_0002603 |
| lateral_to | PATO | http://purl.obolibrary.org/obo/PATO_0001193 |
| located_in | PATO | http://purl.obolibrary.org/obo/PATO_0002261 |
| part_of | RO | http://purl.obolibrary.org/obo/BFO_0000050 |
| posterior_to | PATO | http://purl.obolibrary.org/obo/PATO_0001633 |

# Results and discussion

## Descriptive templates

Semantic-based morphological descriptions were made for 22 species, whose files in RDF/XML format are available at https://github.com/luis-gonzalez-m/*Lepidocyrtus*-RDF-Store. These descriptions have an average of 592 anatomical terms, of which 260 are referred for the chaetotaxy. The RDF triple (see below) expresses (**a**) part-whole relation between anatomical entities (Fig. 2) or (**b**) between anatomical entities and their qualities. The "absences" are represented as (**c**) negations, following the principle of Open Word Assumption (Antoniou and van Harmelen 2008). The importance of this assumption relies on the fact that it is not possible to make a full morphological description of an organism (Wirkner et al. 2017). The term "absence" usually contains evolutive assumptions in a phylogenetic context to explain changes of homologs or appearance of evolutive novelties (Sereno 2007, Willmann 2016), but at descriptive level it is considered as a quality. The above expressions are equivalent to the EQ format and have the logical structure for the character statements (Sereno 2007) and discussed by Göpel and Wirkner 2018.

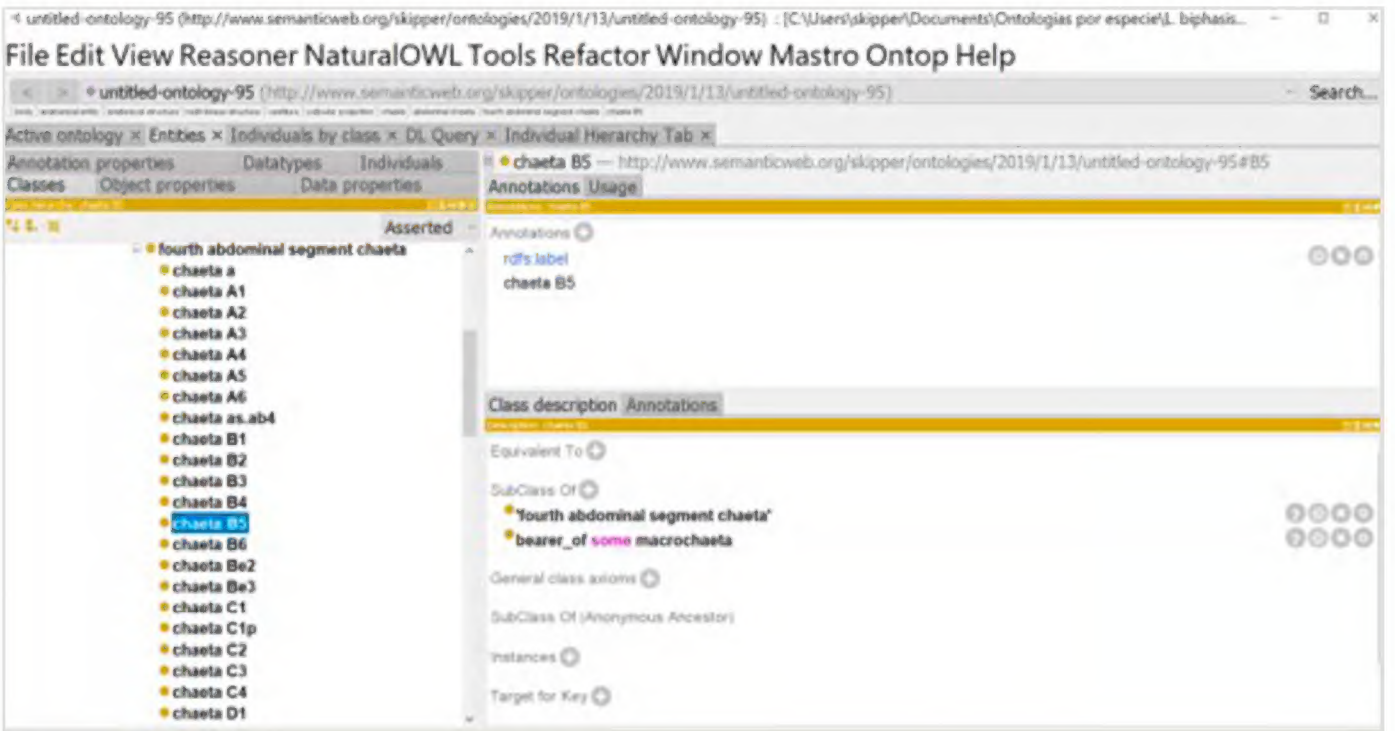


Figure 2. doi

Screenshot of Protégé showing panels used in the class-based method. Left-hand side, some anatomical terms that composed a descriptive template for the species *Lepidocyrtus biphasis* Mari Mutt, 1986. Right-hand side, description for the class “chaeta B5”.

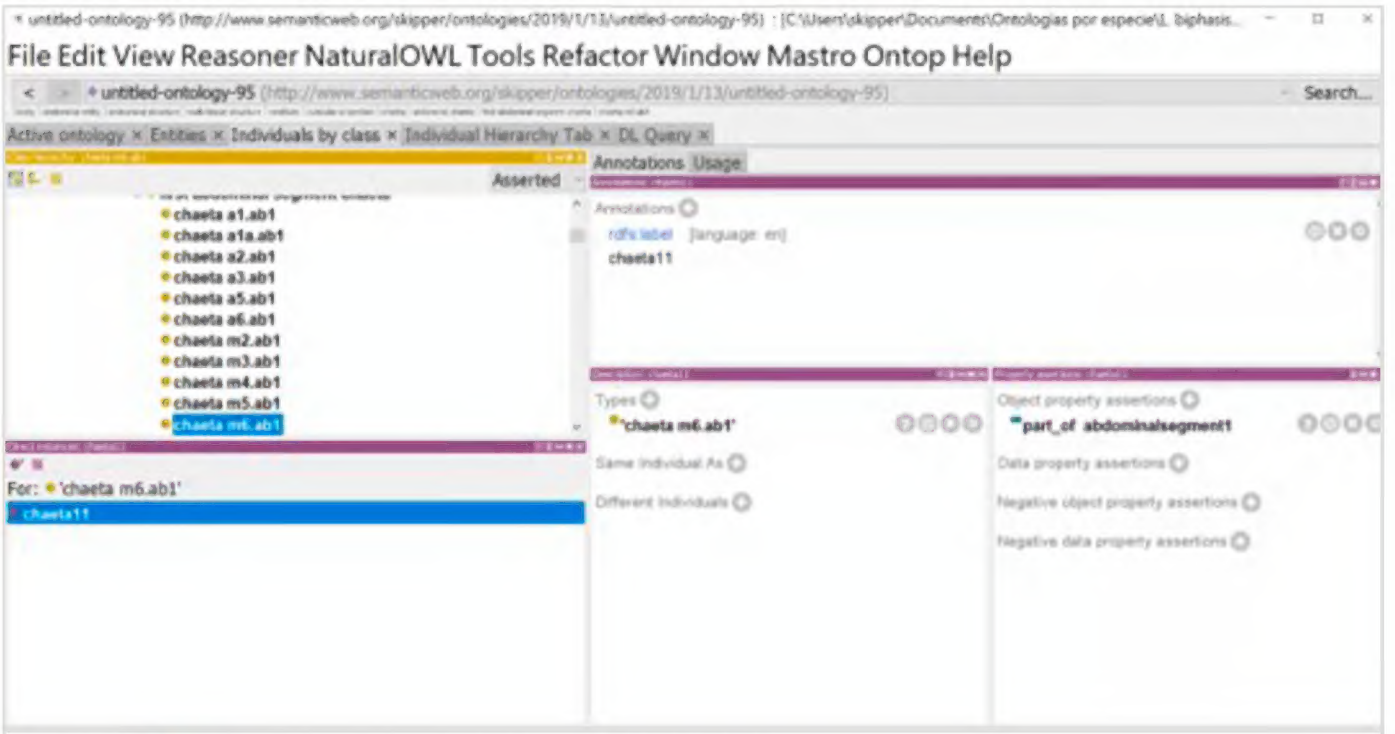


Figure 3. doi

Screenshot of Protégé showing panels used in the class-based method. Left-hand side, the class “chaeta m6.ab1”, the chaeta “m6” located on the abdominal segment 1, and right-hand side, the individual identified by the label “chaeta11” and object property assertion expressed for the species *Lepidocyrtus biphasis* Mari Mutt, 1986.

**chaeta am6.ab3** *part_of* some **abdominal segment 3**.....(**a**)

**chaeta am6.ab3** *bearer_of* some **macrochaeta**.....(**b**)

**chaeta am6.ab3** not (*part_of* some **abdominal segment 3**).....(**c**)

Under the instance-based method, the above is most complex because a second "layer" must be added (Fig. 3), where each class instantiates an individual. For instance, the statement (**d**) is expressed as:

**chaeta145** *part_of* **abdominalsegment3**.....(**d**)

An object property assertion with the provisional labels "chaeta145" and "abdominalsegment3" to name the individuals or parts of organisms perceived in the reality. Some examples of RDF triples and descriptive statements are present in the Table 3.

Table 3.

Examples of RDF triple and descriptive statement.

| RDF triple statement | Descriptive statement (natural language) |
|---|---|
| chaeta A0.h bearer_of some microchaeta | chaeta A0, size: microchaeta |
| chaeta A0.h bearer_of some macrochaeta | chaeta A0, size: macrochaeta |
| chaeta A0.h has_part exactly 1 microchaeta | chaeta A0, number: 1 |
| chaeta Ps4 part_of some cephalic chaeta | chaeta Ps4: present |
| chaeta a3.ab1 part_of some abdominal segment 1 and (anterior_to some chaeta m3.ab1 | chaeta a3, position: anterior to chaeta m3 |
| chaeta a2.ab2 bearer_of some triangular | chaeta a2, shape: triangular |
| chaeta a2.ab3 aligned_with some chaeta m2.ab3 | chaeta a3, alignment: aligned with chaeta m2 |
| chaeta as.ab2 bearer_of some length) and (inheres_in some chaeta m3e.ab2) and (is_approximately_equivalent_to some length) | chaeta as, length: equal to chaeta m3e |
| chaeta D1p bearer_of some smooth | chaeta D1p, texture: smooth |
| dental tubercle bearer_of some domed | dental tubercle, curvature: domed |

## RDF repository

RDF store for biological data is oriented mainly to molecular data with Uniprot (https://uniprot.org) and Bio2RDF (https://bio2rdf.org), while for morphological data, RDF stores have not been developed. When the RDF is available, the next step is the creation of a semantic web service to put the semantic data on the web (Wollbrett et al. 2013). In SPARQL query the clauses WHERE and SELECT permit to extract information stored in RDF stores. The WHERE clause obtains data out of the dataset and SELECT names what parts of the dataset are pulled (DuCharme 2013). These clauses return RDF triples as the answer for the query.

A list of RDF triple, where the subclass relation is retrieved for the chaetae that composed the cephalic chaeta (Table 4) is obtained through the code in the example 1. Likewise, a list

of RDF triple with the displayed relations *bearer_of*, and *posterior_to*, is showed in the Table 5 through the code in the example 2, being a descriptive block for the species *L. americanus* Cipola & Bellini, 2019:

Table 4.
Example 1 (Output)

| ID | SUBJECT | PREDICATE | OBJECT |
|---|---|---|---|
| 1 | chaeta A1.h | subClassOf | cephalic chaeta |
| 2 | chaeta A2.h | subClassOf | cephalic chaeta |
| 3 | chaeta A3.h | subClassOf | cephalic chaeta |
| 4 | chaeta A4.h | subClassOf | cephalic chaeta |
| 5 | chaeta A5.h | subClassOf | cephalic chaeta |

Table 5.
Example 2 (Output)

| ID | SUBJECT | PREDICATE | OBJECT |
|---|---|---|---|
| 1 | chaeta A1 | bearer_of | microchaeta |
| 2 | chaeta A1.h | bearer_of | microchaeta |
| 3 | chaeta A3.h | posterior_to | chaeta A5.h |
| 4 | chaeta a4.lb | bearer_of | smooth |
| 5 | chaeta A5.lb | bearer_of | serrated |

Example 1 (Table 4)

```
PREFIX rdfs:
PREFIX rdf:
PREFIX owl:
SELECT DISTINCT ?s ?p ?o
WHERE { ?s rdfs:subClassOf ?o;
}
```

Example 2 (Table 5)

```
PREFIX rdfs:
PREFIX rdf:
PREFIX owl:
SELECT DISTINCT ?s ?p ?o
WHERE { ?s rdf:type owl:Class ;
   rdfs:subClassOf [ rdf:type owl:Restriction ;
      owl:onProperty ?p ;
      owl:someValuesFrom ?o; ] ;
}
```

chaeta e.lb

The querying of individuals is similar to querying ontology classes in Fuseki, but the main difference is the querying of individuals or objects in named graphs. For example, the SPARQL query to access named graph follows the general syntax (Vogt 2019):

```
SELECT*
WHERE {
GRAPH <123> # 123 is the named graph
}
```

Semantic-based methods for the documenting of morphological data have gained interest in recent years, motivated by the potential application in phylogenetics (Vogt et al. 2009, Balhoff et al. 2013, Tarasov 2019). This application involves the coding and extraction of the character and character states, and recognition of historical and serial homologies across anatomical entities employing computational reasoning (Cui 2012, Mabee et al. 2019). However, the development of computational tools for the generation of phenotypic annotations from RDF triple has received little attention. There are several limitations to its implementation: 1. these methods require datasets such as anatomic ontology, which are not available for most taxonomic groups, 2. these methods are unknown or less preferable than morphological descriptions in natural language, 3. the attention of character statement coding is relevant in comparison with the expression of descriptive statements, which are framed within taxonomic tradition or publication requirements, and 4. technical limitations. In this study, for example, the descriptive statements were generated manually, but other options such as the declaration of RDF triples in spreadsheets and importing to ontology editor may reduce the time-consuming demand. Also, the extraction of phenotypic annotations from texts has important advances (Wood et al. 2003, Thessen et al. 2012, Dahdul et al. 2015, Cui 2012). For instance, the incorporation of Natural Language Processing (NLP) during the extraction of phenotype data sets could increase the efficiency of this task (Dahdul et al. 2015). However, the generation of phenotypic annotations from RDF triple, employing techniques of Natural Language Generation (NLG) has not been explored or applied in morphological descriptions.

The instance-based method, “semantic anatomy instance”, is most complex when the Abox assertions are built with Protégé because individuals by class need to be specified, resulting in a template composed of two layers: ontology class and instanced objects. Recently proto.morphdbase.de incorporates instance-based methods, promising to

increase the use of semantic-based tools during morphological descriptions. This application is complementary to other tools that integrate morphological data with multimedia representations, for instance, MorphoNet (Leggio et al. 2019), MorphoBank (O'Leary and Kaufman 2011), and Phenotools (Eliason et al. 2019). Proto.morphdbase.de has two principal components: semantic graph generated for each organism and accessibility to metadata supplied from scientific publication and regulated by FAIR Guiding Principles to manage stored data (Wilkinson et al. 2016, Vogt 2019).

Currently, there is an imperative need to document biological diversity, which implies the use of computational tools for the processing of different data generated in the biology domain. Unfortunately, this urgency is directed mostly to the storing and processing of molecular data, while the morphology is continually displaced. Morphological descriptions are a useful source of data but due to their nature and complexity requires "creative" solutions, new automatic or semi-automatic methods that permit the interchange between natural language employed commonly in published morphological descriptions and RDF triple syntax. The use of ontologies uncovers the subtle process between morphological data (expressed by RDF triple statement) to character statement, where the character state (properties) arises from the comparison between species and before the building of character matrix.

Initiatives about morphological descriptions that employ standardized languages are not new (Dallwitz 1984, Paterson et al. 2004, Cui 2008). Recently, Cui et al. 2020 developed an author-driven method, where the author or taxonomist defines what ontological classes are necessary for their descriptions. These classes are included by the software engineer in an ontology, this means, the author is immersed in the builiding of the ontology, expressing semantic relations and conflicts in the use of terminology. Although the ability on the management of RDF syntax during phenotypic annotation could be cumbersome, the use of structured syntax in morphological descriptions contributes to making them available to computational analysis.

However, the taxonomic tradition has an important weight in the language employed during morphological descriptions and is taxon-dependent. It is necessary to reconciliate the needs of taxonomists and friendly tools to incorporate these methods. It is not the goal to evaluate the multiple RDF store available, which differs in properties as storage size, querying time, and applicability (Frey et al. 2019). In this study, the building of descriptive templates is suggested, being an intuitive approach to manage RDF triplet and its storage. The semantic-based methods are a tool that increases the management, processing, and comparison of morphological data that could contribute to its integration in the taxonomic work with a friendly computational approach.

## Acknowledgements

The author thanks the financial support of MINCIENCIAS (National doctorate scholarship 727). To Lars Vogt for comments about Semantic Anatomy Instance.

## Funding program

MINCIENCIAS

## Grant title

Scholarship for doctoral studies 727

## Hosting institution

Universidad Nacional de Colombia, Facultad de Ciencias Agrarias.

## Conflicts of interest

None

## References

- Antoniou G, van Harmelen F (2008) A Semantic Web Primer. The MIT Press, United States of America, 287 pp.
- Balhoff J, Dahdul W, Kothari C, Lapp H, Lundberg J, Mabee P, Midford P, Westerfield M, Vision T (2010) Phenex: ontological annotation of phenotypic diversity. Nature Precedings https://doi.org/10.1038/npre.2010.4636.1
- Balhoff J, Kawada R, Mullins P, Deans A (2012) A revision of *Evaniscus* (Hymenoptera, Evaniidae) using ontology-based semantic phenotype annotation. The University of North Carolina at Chapel Hill University Libraries https://doi.org/10.17615/ade1-0906
- Balhoff JP, Mikó I, Yoder MJ, Mullins PL, Deans AR (2013) A semantic model for species description applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia. Systematic Biology 62 (5): 639-59. https://doi.org/10.1093/sysbio/syt028
- Bellinger P, Christiansen K, Janssens F (2021) Collembola. https://collembola.org. Accessed on: 2021-4-19.
- Burks KN, Mikó I, Deans AR (2016) *Dendrocerus mexicali* (Hymenoptera, Ceraphronoidea, Megaspilidae): Novel antennal morphology, first description of female, and expansion of known range into the U.S. ZooKeys 569: 53-69. https://doi.org/10.3897/zookeys.569.6629
- Cui H (2008) Converting taxonomic descriptions to new digital formats. Biodiversity Informatics (5). https://doi.org/10.17161/bi.v5i0.46
- Cui H (2012) CharaParser for fine-grained semantic annotation of organism morphological descriptions. Journal of the American Society for Information Science and Technology 63 (4): 738-754. https://doi.org/10.1002/asi.22618
- Cui H, Dahdul W, Dececchi A, Ibrahim N, Mabee P, Balhoff J, Gopalakrishnan H (2016) CharaParser+EQ: Performance evaluation without gold standard. Proceedings of the

Association for Information Science and Technology 52 (1): 1-10. https://doi.org/10.1002/pra2.2015.145052010020

- Cui H, Zhang L, Ford B, Cheng H, Macklin JA, Reznicek A, Starr J (2020) Measurement recorder: developing a useful tool for making species descriptions that produces computable phenotypes. Database https://doi.org/10.1093/database/baaa079
- Dahdul W, Dececchi TA, Ibrahim N, Lapp H, Mabee P (2015) Moving the mountain: analysis of the effort required to transform comparative anatomy into computable anatomy. Database https://doi.org/10.1093/database/bav040
- Dallwitz MJ (1984) User's guide to the DELTA system. A general system for coding taxonomic descriptions. 2nd ed. Canberra, Division of Entomology, CSIRO https://doi.org/10.25919/wmhe-qc95
- De Giacomo G, Lenzerini M (1996) Tbox and Abox reasoning in expressive description logics. AAAI. URL: www.aaai.org
- DuCharme B (2013) Learning SPARQL: querying and updating with SPARQL 1.1. O'Reilly Media, Inc.
- Eliason C, Edwards S, Clarke J (2019) Phenotools: an R package for visualizing and analyzing phenomic datasets. Methods in Ecology and Evolution https://doi.org/10.1111/2041-210X.13217
- Frey J, Müller K, Hellmann S, Rahm E, Vidal M (2019) Evaluation of metadata representations in RDF stores. Semantic Web 10 (2): 205-229. https://doi.org/10.3233/sw-180307
- Girón J, Mikó I, González-Montaña LA, Tarasov S, Matentzoglu N (2021) Ontology for the Anatomy of the Insect SkeletoMuscular system (AISM). URL: https://doi.org/10.5281/zenodo.4660322
- Gonzalez-Montaña L (2021) Collembola anatomy ontology (CLAO). Zenodo https://doi.org/10.5281/zenodo.4660386
- Göpel T, Wirkner CS (2018) Morphological description, character conceptualization and the reconstruction of ancestral states exemplified by the evolution of arthropod hearts. PLOS One 13 (9): e0201702. https://doi.org/10.1371/journal.pone.0201702
- Grumbling G, Strelets V (2006) FlyBase: anatomical data, images and queries. Nucleic Acids Research 34 (Database issue): 484-8. https://doi.org/10.1093/nar/gkj068
- Leggio B, Laussu J, Carlier A, Godin C, Lemaire P, Faure E (2019) MorphoNet: an interactive online morphological browser to explore complex multi-scale data. Nature Communications 10 (1): 2812. https://doi.org/10.1038/s41467-019-10668-1
- Mabee PM, Balhoff JP, Dahdul WM, Lapp H, Mungall CJ, Vision TJ (2019) A Logical Model of Homology for Comparative Biology. Systematic Biology 69 (2): 345-362. https://doi.org/10.1093/sysbio/syz067
- Mahner M, Bunge M (1997) Foundations of biophilosophy. Springer- Verlag Berlin Heidelberg, Berlin, 423 pp. https://doi.org/10.1007/978-3-662-03368-5
- Mari Mutt J, Bellinger P (1990) A Catalog of the Neotropical Collembola, Including Nearctic Areas of Mexico, Florida. Sandhill Crane Press
- Mikó I, Copeland R, Balhoff J, Yoder M, Deans A (2015) Folding wings like a cockroach: a review of transverse wing folding ensign wasps (Hymenoptera: Evaniidae: *Afrevania* and *Trissevania*). PLOS One https://doi.org/10.1371/journal.pone.0094056
- Mikó I, Trietsch C, Sandall E, Yoder MJ, Hines H, Deans AR (2016) *Malagasy Conostigmus* (Hymenoptera: Ceraphronoidea) and the secret of scutes. PeerJ 4 https://doi.org/10.7717/peerj.2682

- Musen M (2015) The protégé project. AI Matters 1 (4): 4-12. https://doi.org/10.1145/2757001.2757003
- O'Leary M, Kaufman S (2011) MorphoBank: phylophenomics in the "cloud". Cladistics 27 (5): 529-537. https://doi.org/10.1111/j.1096-0031.2011.00355.x
- Paterson T, Kennedy J, Pullan M, Cannon A, Armstrong K, Watson M, Raguenaud C, McDonald S, Russell G (2004) A universal character model and ontology of defined terms for taxonomic description. Lecture Notes in Computer Science 63-78. https://doi.org/10.1007/978-3-540-24745-6_5
- Ramírez M, Michalik P (2019) The Spider Anatomy Ontology (SPD)—A versatile tool to link anatomy with cross-disciplinary data. Diversity 11 (10). https://doi.org/10.3390/d11100202
- Sereno P (2007) Logical basis for morphological characters in phylogenetics. Cladistics 23 (6): 565-587. https://doi.org/10.1111/j.1096-0031.2007.00161.x
- Silva TSR, Feitosa RM (2019) Using controlled vocabularies in anatomical terminology: A case study with Strumigenys (Hymenoptera: Formicidae). Arthropod structure & development 52: 100877. https://doi.org/10.1016/j.asd.2019.100877
- Tarasov S (2019) Integration of anatomy ontologies and evo-devo using structured markov models suggests a new framework for modeling discrete phenotypic traits. Systematic Biology 68 (5): 698-716. https://doi.org/10.1093/sysbio/syz005
- Thessen AE, Cui H, Mozzherin D (2012) Applications of natural language processing in biodiversity science. Advances in Bioinformatics 2012: 391574. https://doi.org/10.1155/2012/391574
- Topalis P, Tzavlaki C, Vestaki K, Dialynas E, Sonenshine DE, Butler R, Bruggner RV, Stinson EO, Collins FH, Louis C (2008) Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. Insect Molecular Biology 17 (1): 87-9. https://doi.org/10.1111/j.1365-2583.2008.00781.x
- Trietsch C, Mikó I, Notton D, Deans A (2018) Figure 2 from: Trietsch C, Mikó I, Notton D, Deans A (2018) Unique extrication structure in a new megaspilid, *Dendrocerus scutellaris* Trietsch & Mikó (Hymenoptera: Megaspilidae). Biodiversity Data Journal 6: e22676. https://doi.org/10.3897/BDJ.6.e22676. Biodiversity Data Journal https://doi.org/10.3897/bdj.6.e22676.figure2
- Vogt L, Bartolomaeus T, Giribet G (2009) The linguistic problem of morphology: structure versus homology and the standardization of morphological data. Cladistics 26 (3): 301-325. https://doi.org/10.1111/j.1096-0031.2009.00286.x
- Vogt L (2017) Assessing similarity: on homology, characters and the need for a semantic approach to non-evolutionary comparative homology. Cladistics https://doi.org/10.1111/cla.12179
- Vogt L (2019) Morphological descriptions in times of eScience: instance-based versus class-based semantic representations of natomy. Unpublished https://doi.org/10.13140/rg.2.2.28314.29124
- Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Santos L, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray A, Growth P, Goble C, Grethe J, Heringa J, Hoen P, Hooft R, Kuhn T, Kok J, Lusher J, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz G, Thompson M, Lei J, Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR

Guiding Principles for scientific data management and stewardship. Scientific data 3 (1). https://doi.org/10.1038/sdata.2016.18

- Willmann R (2016) The evolution of Willi Hennig's phylogenetic considerations. The Future of Phylogenetic Systematics128-199. https://doi.org/10.1017/cbo9781316338797.009
- Wirkner C, Göpel T, Runge J, Keiler J, Klussmann-Fricke B, Huckstorf K, Scholz S, Mikó I, Yoder M, Richter S (2017) The first organ-based free ontology for arthropods (Ontology of Arthropod Circulatory Systems - OArCS) and its integration into a novel formalization scheme for morphological descriptions. Systematic Biology https://doi.org/10.1093/sysbio/syw108
- Wollbrett J, Larmande P, de Lamotte F, Ruiz M (2013) Clever generation of rich SPARQL queries from annotated relational schema: application to Semantic Web Service creation for biological databases. BMC Bioinformatics 14 (1). https://doi.org/10.1186/1471-2105-14-126
- Wood M, Lydon S, Tablan V, Maynard D, Cunningham H (2003) Using parallel texts to improve recall in IE. Proceedings of Recent Advances in Natural Language Processing (RANLP '03).
- Yoder M, Mikó I, Seltmann KC, Bertone MA, Deans AR (2010) A gross anatomy ontology for Hymenoptera. PLOS One 5 (12): e15991. https://doi.org/10.1371/journal.pone.0015991
- Yoder M, Twidale M, Thomer A, Vogt L, Franz N, Guo J, Deans A, Balhoff J (2018) Taxonomy and the production of semantic phenotypes. In: Thessen A (Ed.) Application of semantic technology in biodiversity science.